The Businesses of Open Data and Open Source: Some Key Similarities and Differences

Juho Lindman and Linus Nyman

"It's difficult to imagine the power that you're going to have" when so many different sorts of data are available.

Tim Berners-Lee
Inventor of the World Wide Web

Open data and open source are phenomena that are often automatically grouped together, perhaps because they share the word "open". A careful analysis of what open means in each of these cases is a stepping stone towards building viable businesses around both open source applications and on open data. Although there are, indeed, elements they share through their openness, the ways in which they differ are significant. In this conceptual paper, we aim to outline the differences and similarities of the two phenomena from a commercial perspective.

Introduction

Open source and open data both have a focus on "openness", and most developers and researchers could easily identify similarities between the two phenomena. For example, both open source and open data are enabled – or at the very least greatly helped – by the Internet, which provides a backbone for collaborative development efforts, communication infrastructure, as well as a means to support the sharing of both application and data. However, open source and open data are distinct phenomena with significant differences, and these differences clearly impact how commercial success can be achieved in each domain.

Given that TIM Review readers tend to be more familiar with open source than open data, our goal is to explore the concept of open data through a comparison with open source and with an emphasis on the similarities and differences that are relevant to technology businesses. We focus on three key questions:

1. Are the phenomena similar?

2. Are the licenses of software and data similar?

3. Are the businesses and revenue models similar?

Understanding these two phenomena is useful to managers and entrepreneurs interested in the business po-

tential of the released data sets. This understanding is also useful for open data proponents who are interested in the business aspects of open source and the lessons that the business of open source offers. Designers of related services may be interested what potential open data and open source have to offer in terms of novel and better service opportunities.

The structure of the article is as follows. We first describe key characteristics of open source and open data. We then compare these two phenomena from three business-oriented perspectives: licensing, commercial aspects, and relevant actors. Finally, we provide some takeaways for managers and entrepreneurs.

Comparing Open Source and Open Data

In computer science, in theory as well as in practice, the distinction between data and application is critical. Therefore, the most obvious and fundamental difference between open data and open source is that the former focuses on the data and the latter focuses on applications.

Data has multiple meanings, including any end-product of measurement, but in this investigation, we use a slightly more technical definition of data: data refers to stored symbols. Data is considered a resource – raw material for the application. Open data means data that is technically and legally made available for re-

The Businesses of Open Data and Open Source: Some Key Similarities and Differences
Juho Lindman and Linus Nyman

use and republication. The underlying idea is that the increased transparency will help to create trust in users and developers, as well as offer a way to create new services based on the collected data. In many cases, the data is collected by government entities for various purposes and thus additional economic value would be created when the published data is put to use. However, open data includes open government-collected data as well as data released by private actors.

An application, on the other hand, is compiled source code that operates on data. Open source refers to a legal and technical arrangement related to software production that results in open source code that is accepted under an license that complies with the Open Source Definition (opensource.org/definition). These licenses are based on the copyright protection of the code; thus, the "open" of open source refers to the source code.

To summarize: the first significant difference between open data and open source is that of data versus application. Data can be numbers, locations, names, etc. In and of itself, data does nothing. Source code, or rather an application, is something that uses or produces data. These two aspects, although they rely on one another for their significance, are different in both essence and purpose. Indeed, it is some of these differences that this article seeks to point out and clarify.

Comparing Licensing Aspects

The key similarity between open data and open source lies in the prerequisite of openness. But what, exactly, is it that is open, and are there degrees, or types, of openness? For open source software, the openness primarily means a guaranteed access to the application's source code as well as an arrangement that makes sure that the code can be forked, modified, and redistributed. (For more on the significance of the right to fork, see Nyman and Lindman [2013; timreview.ca/article/644].)

For open data, a similar "access principle" provides access to the data (and metadata) and it provides the opportunity to reuse it in applications. Data also needs to be maintained and updated. The actor that collects or mashes up the data from different sources usually has the option to stop providing access or maintenance to the data (i.e., to "close" the data).

Source code can be copyrighted (or copylefted: tinyurl.com/qygb2) but, in some cases, data falls outside copyright protection. Whether particular data can be copyrighted is subject to national legislation. However, copyright is not the only law that applies to data; depending on its content, other laws may also regulate its use. Laws may govern the collection, storage, maintenance, access, use, and representation of data. For example, laws relating to sensitivity, privacy concerns, or national security may apply to different datasets.

Table 1 compares different licensing aspects of open source versus open data. The legal arrangements (i.e., copyright, licenses, original publisher, and role of contracts) for the two phenomena are different. For a more thorough discussion on data licensing, see the "Guide to Open Data Licensing" (tinyurl.com/lkhg6df), which is published by the Open Definition project.

Table 1. Licensing of open data versus open source

	Open Source (Application)	Open Data (Data)
Copyright	Applies to all source code	May apply
Licensing	Licenses must comply with the Open Source Definition (opensource.org/osd-annotated)	Relevant if protected by copyright. Possible licensing options include the Wikimedia commons (commons.wikimedia.org) and the Creative Commons (creativecommons.org)
Original publisher	Several versions (distributions) of application and forking are possible	Data often collected, maintained, and controlled by data publisher
Contracts	Normally not required; the license agreement defines the rights of the developers and users	Data publisher may have an incentive to monitor data use or to create feedback loops to reusers of their data

The Businesses of Open Data and Open Source: Some Key Similarities and Differences
Juho Lindman and Linus Nyman

Comparing Commercial Aspects

The business of open source is in itself a diverse field, with companies generating income from various sources related to open source products. The two main categories of income are those from support and services related to a program (e.g., training, consulting, feature development) and from selling closed source versions of an open source program, a practice called "dual licensing". For a primer on the business of open source, see Widenius and Nyman (2014) in this issue of the TIM Review.

In contrast, the business of open data is a young field, but it holds promise for service innovation. Discussion on the viable revenue sources is still ongoing. The data publisher or the user of the application pays the costs related to the collection, maintenance, and enrichment of data, but customers normally do not pay subscription fees for accessing data.

In the following subsections, we compare the expense-related and income-related considerations of businesses that rely on open data or open source .

Who pays the bill, and why?
Open source can save firms money if they are able to attract free community participation. However, companies may also be willing to support open source development, for example by paying a developing group or foundation, or by assigning its own developers to an open source project. Even though anyone, even the firm's competitors, can then benefit from any improvements to the code, this approach is common in open source development. Typically, firms use this strategy to develop aspects of their product offering that would be considered "table stakes" (tinyurl.com/5u4aut). By collaborating – even with competitors in the same or similar markets – to develop non-unique

foundational aspects of an application, companies save time and development effort, which can then be redirected into developing the aspects of their offering that will differentiate them from their competitors.

In addition to the costs of collecting data, open data providers must often spend money and effort both to clean up the data for publishing as well as for keeping it open. With such tasks, providers may benefit from community participation, just as in the case of open source software. Issues related to "community management" are therefore similar in both cases.

Who makes money, and how?
Openness usually means that an application or dataset can be acquired free of charge. In the case of open data, the publishers are normally considered to have given permission for others to build services on top of their released dataset. The services provided by these other parties may add value above and beyond just the provision of data, and the costs of designing the applications, collecting the data, and maintaining the services are covered by various different arrangements depending on the motivations of the other parties, their possible business models, and the nature of the value created.

Value can stand for both economic value (i.e., money) or a wider benefit. The openness of both open source and open data can potentially offer either one of these two types of value. However, it is notable that the dynamics that produce value are different. In Table 2, we list some of the benefits perceived by the key actors in each case. The table is not exhaustive list, but it provides an illustration of topical issues.

One of the main differences in the business aspects of open data and open source is that, at least currently, it is rare for the data provider to make money on open

Table 2. Examples of key value sources in open source versus open data

	Actor	Economic value	Other value
Open Source	Companies	Dual licensing, support and services	Product innovation, platform innovation
	Customers	Cost savings	Evade vendor lock-in
	Actor	**Economic value**	**Other value**
Open Data	Data owner	Sales of premium access	Public service, receive additional developmental resources
	3rd party	Sell applications	Increased transparency, novel services

The Businesses of Open Data and Open Source: Some Key Similarities and Differences

Juho Lindman and Linus Nyman

data. The main perceived sources of benefits are related to public service or to situations where the data needs to be collected and maintained anyway. Releasing the datasets would then enable others to benefit from them and may result in new and useful services. Typically, the funding for the collection and maintenance of the data in such situations also comes from public sources. Normally, a company waives the possibility of data sales when it decides to release a dataset. However, we speculate that open data could also be accompanied by a "premium access" option, meaning access to more real-time data, faster access, or access to datasets that contain both open and closed data.

For open source, the commercial actors have established business models that rely on, for example, dual-licensing. Open source has tried and true ways to cut costs and evade lock-in situations. However, many open source contributions are driven not solely by commercial interests, but by the desire for useful software that addresses specific needs, among other motives.

Open data can offer opportunities for downstream service provision. In such situations, some actors that provide open data might be keen to share their costs with the downstream actors that make a profit. It is possible to sell downstream applications, such as closed source software subscriptions. Developers might also have other motivations in writing software that uses open data, such as increased transparency, new visualizations, service provision, etc.

Comparing Elements and Actors

In comparing the elements and actors involved in open data versus open source, we limit our investigation to the four questions illustrated in Figure 1:

1. Who are the main actors?
2. How is the output developed?
3. What is the output of the development?
4. Who is interested in the output?

In an open source project, both a corporate community and an open source community can participate in the software development. When data is being developed for release, data consultants as well as those who clean data participate to the process. Output of the processes are released as open source programs and as published open data.

As shown in Figure 1, the main difference concerning the processes is that the open source process is more open than the open data process. The developers are able to participate in open source software development with varying motivations. For community driven projects in particular, motivations commonly are not financial. In open data, the data publisher is usually expected to carry costs related to releasing the data, such as the costs of collection, aggregation, and anonymization. If these services are outsourced, there is new business opportunity for companies that provide them.

The output is also different: in open data, the data ultimately remains the same through the process, whereas the open source development process aims to change the software. The software is an end in itself, whereas the released open dataset is just the first step in providing the service to the customer.

Conclusions

Although open source licensing has established its value for developers in enabling a viable development model, the business of open data needs further study. Nonetheless, both open source and open data hold potential for business. Their main difference as phenomena is that of data versus application.

Proponents of both phenomena promote the openness of the output, which offers transparency but also changes the competition dynamic. The open source alternative hampers traditional software subscription sales. Open datasets can be easily copied, but the original data collector still has a prominent role in the maintenance of the said dataset: a copied dataset, if not maintained properly, may soon become obsolete.

The commercial potential of open source has been tested and proven over the years, and several business models have emerged. However, the business of open data is still in a pioneering phase. The publisher's role is critical in any open data business, but the publisher might not be the actor that benefits most from data publication; the greater opportunity for entrepreneurship may lie downstream, where value may be captured from services built upon the open data.

Despite their differences, both open source and open data aim to attract development efforts far beyond the originator of the project. Contributors may be driven by a variety of motivations, not excluding economic gain. Economic gains are feasible and attainable, but capturing them requires entrepreneurs and managers to understand the differences in the development process and economic value capture.

The Businesses of Open Data and Open Source: Some Key Similarities and Differences

Juho Lindman and Linus Nyman

Open Source
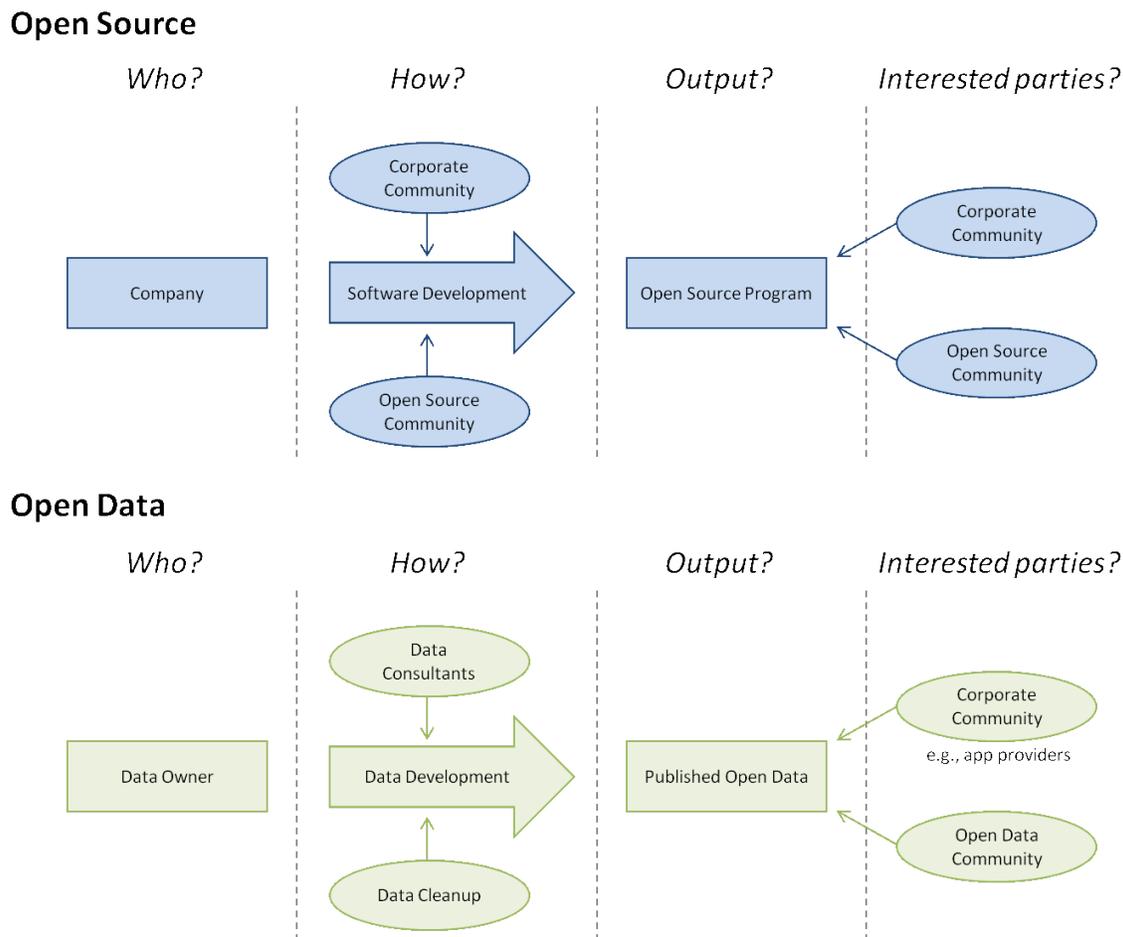
Open Data

Figure 1. Comparing actors and activities of an open data versus open source project

Insights for managers and entrepreneurs
When evaluating business models based on open data, managers and entrepreneurs should consider the following key questions:

1. Do you have sufficient familiarity about the dynamics of open data ecosystems and the required technical capabilities? Open data is a significantly different field from open source; in-depth knowledge of one does not automatically guarantee sufficient knowledge of the other, although open source experts will easily find similarities.

2. If the data is not yours, how certain are you that it will continue to be provided openly in the future? How can you safeguard the relationship between the data collector/maintainer?

3. What is the legal status of the data? Does it allow fees, and what would happen if fees were introduced?

4. What is the license of the software application? Is it possible to gain software subscription revenue from a proprietary application built on the open data stack?

5. What national and international legislation poses obstacles to the service? Whereas software is relatively free of these concerns, many datasets may contain sensitive information. Open datasets can also be combined with other (also private) datasets, and this combination may raise new legal issues.

6. How will you attract developers? What are the dynamics of the development community? How will you support a relationship that takes into account different motivations to participate and benefits the different actors?

In conclusion, a final similarity we can note regarding open source and open data is that open data is now in a

The Businesses of Open Data and Open Source: Some Key Similarities and Differences

Juho Lindman and Linus Nyman

position not dissimilar to that of open source some two decades ago: a new, interesting phenomenon with promise, but also skepticism and challenges regarding an entrepreneurial potential for revenue creation and value capture.

The business models and strategies surrounding open source did not evolve overnight. Open data is an emerging field that may have many opportunities yet to be discovered. We believe that the world of open data holds great untapped potential for knowledgeable entrepreneurs that can identify opportunities for its use.

About the Authors

Juho Lindman is an Assistant Professor of Information Systems Science at the Hanken School of Economics in Helsinki, Finland. Juho's doctoral dissertation from the Aalto University School of Economics in Helsinki focused on open source software development organization In the field of information systems, his current research is focused in the areas of open source software development, open data, and organizational change.

Linus Nyman is a doctoral researcher at the Hanken School of Economics in Helsinki, Finland, where he is researching code forking in open source software. A further research interest of his is free-to-play gaming. He also lectures on corporate strategy, open source software, and the new business models of the Internet age. Linus has a Master's degree in economics from the Hanken School of Economics.